

Final Report on the Analysis of STARS Data

**Prepared by the Center for Evaluation and Assessment
University of Nevada, Las Vegas**

**Gregory Schraw, Ph.D., M. Stat., Co-Director
Keith Zvoch, Ph.D.**

June 10, 2004

Final Report on the Analysis of STARS Data

Executive Summary

We investigated four research questions related to the relationship between state developed criterion-referenced assessments (CRT: STARS) and norm-referenced assessments (NRT), and the extent to which a variety of student and teacher variables relate to district proficiency and compliance with state and federal standards. STARS and NRT assessments were moderately to highly correlated within and across student cohorts (i.e., within and across academic years and grades). Cohort performance on the NRT assessments was stable over the 2001-2003 timeframe; however, the 2003 cohort showed significant proficiency gains over the 2001 cohort on the STARS assessments among 4th, 8th, and 11th grade students. Student and teacher variables explained a significant, but modest proportion of the district-level variation in proficiency and compliance rates. Three district characteristics, student attendance rates, teacher experience, and teacher education, were most predictive of district performance. Other district characteristics, including mobility and graduation rates, district size, and student demographics did not improve prediction. We conclude that STARS provides a more sensitive measure of learning progress. We also suggest that the NDE collect data that is more closely tied to classroom practice and procedure in order to better understand why some districts are more able to reach targeted proficiency levels and comply with state and federal accountability standards.

Background and Purpose

The purpose of this report is to document STARS and NRT assessment results using data from the 2000-2001 and 2002-2003 academic years. STARS (School-based teacher-led assessment and reporting system) places special responsibility on each school district to create local criterion referenced tests (CRTs) in reading and writing. The goal of STARS is to tailor assessments to local curricular goals in order to better align instruction and assessment. STARS requires districts to set learning goals and develop assessments to ensure that learning goals are met. The intent of STARS is to increase the reliability and validity of accountability decisions (Marion,

White, Carlson, Erpenbach, Rabinowitz, & Sheinker, 2002). STARS also is intended to complement norm-referenced tests (NRTs) such as the Iowa Test of Basic Skills (ITBS). The main goal of the analyses reported here is to evaluate and assess relationships between the locally-constructed STARS assessments and the national NRT assessments and to determine whether district characteristics are predictive of student performance on these measures.

It is important to note four caveats in relation to interpreting these results. First, STARS assessments may differ from district to district depending on the learning goals established by each district. Comparisons among districts should be interpreted with caution and causal comparisons among districts are unwarranted. Second, results of these analyses may be affected by district size. Larger districts yield data that is more stable than that associated with smaller districts. Third, some districts did not report data, which may affect the generalizability of the results. Fourth, results presented here are based on the “school district” as the unit of analysis. A district average may not be representative of all schools within that district.

Research Questions

Four research questions were identified in consultation with Dr. Pat Roschewski. These questions align with goals 3 and 4 of the master plan to validate Nebraska’s standards, assessment, and accountability system. These questions focus on data collected in 2001 and 2003. Comparisons with 2003-2004 data will be considered in a separate proposal.

Question 1 addresses the relationship between STARS assessments and NRTs collected concurrently on students in the 2001 and 2003 school years. The main purpose of this analysis was to summarize performance on each type of measure and examine the relationship between CRTs and NRTs within and between the two-targeted years.

Question 2 addresses change in two different cohort groups from 2001 and 2003. By cohort, we mean a group of students who progress through the educational system together. Students from the 2000-2001 academic year represent a different cohort than students from the 2002-2003 academic year. Differences between these cohorts could be due to instruction or individual differences of students. Because cohorts are naturally occurring, rather than

randomly assigned, differences across cohorts are suggestive rather causal in nature. Nevertheless, cohort analyses enable us to draw descriptive conclusions about progress within districts and across the state. The main purpose of these analyses was to compare change among the two cohort groups using STARS and NRT data.

Question 3 addresses the relationship between state writing data and STARS reading data for grade 8 students in the 2002-2003 school year. One purpose of this analysis was to determine whether writing and reading scores are correlated across districts. A second purpose was to investigate which variables are most predictive of writing and reading success at the district level. We used two main types of variables, including student demographic information such as percent ELL and percent free and reduced lunch (FRL), as well as teacher demographics such as annual salary and percent of teachers with graduate degrees.

Question 4 addresses the issue of district compliance with state and federal accountability standards. The purpose of this analysis was to identify factors that discriminate between districts that are in compliance with state and federal (No Child Left Behind) standards and those that are not. A variety of student and teacher variables were used to determine which of these variables, or what combination of these variables, best explains inclusion in the compliance versus no-compliance groups. This information may help the NDE and individual districts focus on variables of special importance with respect to reaching state and federal accountability criteria.

Methods and Procedures

The state of Nebraska has 517 school districts. Accountability data on each district is posted on the NDE website. However, to protect student and teacher anonymity, the NDE suppresses data from some of the smaller school districts. Our analyses focused on districts with a minimum of 50 students to allow for dissemination of accountability data. Typically, our analyses included between 200 and 300 districts.

A master database was compiled using over 20 separate of data files available on the NDE website. Individual data files usually had some missing data unrelated to district size. For example, approximately 10% of districts did not report average teacher salary. The cause of missing data is

unknown to the authors of this report. In addition, it is unclear whether missing data affects in any significant way our interpretations or the generalizability of the results (Little & Rubin, 1987; Schafer, 1997). Replication across different cohorts would help address this issue.

Variables used in the analyses were screened prior to analyses. Most variables were distributed normally. When a variable was skewed and/or kurtotic, the variable was trimmed by excluding extreme scores unless otherwise noted. For example, when comparing STARS reading and writing data, a disproportional number of districts reported that 0% of students reached proficiency. These districts tended to be small in size, usually with less than 50 students. We excluded schools with less than 50 students in this case.

Several aspects of the master database and analytic procedures should be noted.

1. STARS reading data were analyzed in terms of the aggregate total score reported by NDE. The total score was used for two reasons. The greater reliability of the total score (relative to individual standard scores) was the first reason. The second was to reduce the number of analyses reported to a manageable size.
2. Separate analyses were conducted for 4th, 8th, and 11th grades. This level of analysis allowed us to examine the percentage of students reaching proficiency at each of the three tested levels (and/or compliance with state and federal standards).
3. There was no STARS writing data for the 2000-2001 academic year, so a cohort writing comparison could not be conducted.
4. The English Language Learner (ELL) variable tended to be highly skewed, with most districts reporting very few ELL students. We included this variable in our analyses *without deleting 0% districts* (i.e., districts without ELL students). This may lead to a conservative estimate of the effect of ELL students due to restriction of range among scores.

Results

A number of different analyses were conducted. The general purpose and nature of the statistical procedure for each research question is described below. An attempt was made to reduce large amounts of data into concise

summary tables. Complete data files and statistical analyses are available upon request.

Question 1. We examined the mean percentage of students meeting proficiency for 4th, 8th, and 11th grade for STARS, math NRT, and reading NRT scores for the 2001 and 2003 academic years. We also computed the correlation across the 2001-2003 years for each variable, as well as computing the correlation between math and reading within each cohort.

Table 1 shows the mean percentage of students who reached proficiency on six different performance measures. In most cases, 50% to 65% of students reached proficiency. Proficiency levels were higher for grade 11 than grades 4 and 8. One possible reason is that assessments are more specific at grade 11 and may be better aligned to standards. It should also be noted that whereas the 2003 student cohort was more proficient than the 2001 cohort on the STARS assessment, the reverse was generally true of cohort performance for the NRT measures.

Table 1: Means and Standard Deviations for 2001 and 2003 Percent Proficient for STARS and NRT Data

	4thGrade		8 th Grade		11 th Grade	
Type of Performance Measure	Mean	SD	Mean	SD	Mean	SD
Percent STARS Proficient 2001	.59 N = 306	.32	.48 N = 382	.37	.73 N = 250	.17
Percent STARS Proficient 2003	.63 N = 281	.33	.56 N = 326	.34	.74 N = 242	.13
Percent NRT Reading 2001	.64 N = 209	.14	.62 N = 201	.13	.58 N = 174	.14
Percent NRT Reading 2003	.52 N = 263	.29	.47 N = 305	.30	.59 N = 203	.16
Percent NRT Math 2001	.66 N = 204	.16	.66 N = 197	.16	.66 N = 173	.14
Percent NRT Math 2003	.53 N = 260	.31	.51 N = 303	.33	.65 N = 205	.16

Note: N equals number of school districts.

Table 2 shows the correlations between 2001 and 2003 district performance measures. Rows 1-3 of Table 2 indicate whether the two cohorts have similar performance over time. In general, the 2001 and 2003 cohorts perform similarly in 4th and 8th grades, but do not perform similarly among 11th grade students (i.e., with the exception 11th graders, districts that have high (or low) levels of proficiency in 2001 have similar rank proficiency levels in 2003). The lower 11th grade correlations may reflect differences (or a change) in either instruction and/or assessment at the high school level.

The data in rows 1-3 also indicate that the 2001-2003 cohort correlations tend to be much larger for STARS assessments compared to reading and math NRT scores. This may be due to better alignment between instruction and STARS assessments at the district level.

Finally, rows 4-5 of Table 2 indicate that math and reading NRT performance is highly correlated within a cohort group. Those who do well in math tend to do well in reading. This relationship is strong in the 2001 cohort and very strong in the 2003 cohort. The change between 2001 and 2003 may be due to better alignment of instruction and assessment as districts develop and implement STARS.

Table 2: Correlations Between 2001 and 2003 STARS and NRT Data

	4thGrade	8 th Grade	11 th Grade
STARS 2001-2003 Proficiency	.75	.88	.23
NRT 2001-2003 Reading Proficiency	.55	.41	.40
NRT 2001-2003 Math Proficiency	.47	.45	.30
NRT 2001 Math and 2001 Reading	.61	.63	.62
NRT 2003 Math and 2003 Reading	.91	.95	.62

Note: All correlations are significant at the $p < .01$ level.

Question 2. We examined the mean change in the percentage of students meeting proficiency for 4th, 8th, and 11th grade for STARS, reading NRT, and math NRT scores for the 2001 and 2003 academic years. The purpose of these analyses was to examine whether there is a systematic change in cohort performance over time.

The No Child Left Behind (NCLB) federal legislation mandates that schools demonstrate adequate yearly progress (NCLB: No Child Left Behind, 2002).

For NCLB to be met, schools must demonstrate a higher rate of proficiency across successive student cohorts . Thus, NCLB expects the 2003 cohort to have a higher percentage of students reaching proficiency than a 2001 cohort. We computed change scores (i.e., a value of .05 represent a 5% increase in the percentage of students reaching proficiency, whereas a score of -.05 indicates a 5% decrease) and conducted one-tailed significance tests on the change scores to determine whether proficiency rates increased from 2001 to 2003.

Table 3 reveals two different results. The first is that the level of proficiency increased significantly between cohorts when using STARS assessments. The second result indicates that reading and math NRT proficiency levels did not change over the same time period. There are two plausible explanations for this pattern. The first is that the increase in STARS data is due to chance factors. A second explanation is that STARS assessments are better aligned to local district instruction and therefore yield more sensitive measures of progress. While it is impossible to disentangle the effects of chance and better alignment between instruction and assessment, it seems reasonable to assume that local assessments characteristic of STARS would be better able to detect change.

It is important to note that yearly progress shown in Table 3 is small. NCLB mandates progress of 3% to 5% a year on average. In contrast, Linn (2000, 2003) reports a general trend of 1% progress per year. The STARS progress shown in Table 3 is consistent with the trends reported by Linn (2003).

It should be noted that the difference scores shown in Table 3 are based on districts reporting data for both 2001 and 2003; thus, the number of districts is smaller than in Table 1. It is possible that complete data from all districts might have changed the results reported in Table 3.

Table 3: Means, Standard Deviations, and t Tests for Change in Percent Proficient Scores

	4 th Grade	8 th Grade	11 th Grade
	Mean SD	Mean SD	Mean SD
STARS Percent Change 2001-2003	.050 .15 N = 218 t = 4.18 p < .01	.018 .16 N = 288 t = 1.87 p < .05	.019 .18 N = 228 t = 1.65 p < .05
NRT Reading Percent Change 2001-2003	.003 .14 N = 173 t = .36 p > .70	-.001 .16 N = 180 t = -.10 p > .90	.014 .17 N = 154 t = 1.06 p > .25
NRT Math Percent Change 2001-2003	.008 .18 N = 167 t = .62 p > .50	.007 .16 N = 174 t = .59 p > .50	-.002 .17 N = 154 t = -.18 p > .80

Note: N equals number of school districts. Significance of t values are based on one-tail tests.

Question 3. We examined whether student and teacher variables were significant predictors of the percent of students reaching proficiency on STARS reading and writing performance measures using data from 8th grade students. Tables 4 and 5 list a number of predictor variables which are described in detail on the NDE website. We conducted two regression analyses; one using 2003 STARS reading data, and the second analysis using 2003 STARS writing data. The correlation between percentage of students reaching reading and writing proficiency was .326. Although this correlation is statistically significant, it is somewhat low in our opinion. It suggests that there is a small degree of overlap in reading and writing proficiency at the district level.

The data in Table 4 is based on 219 districts. All variables were entered into the regression equation at the same time to predict the percentage of students reaching proficiency on the STARS reading assessment. Tables 4 and 5 show the mean percentage, standard deviation, beta weight (i.e., the relative contribution of the variable in the prediction equation), t value, p value (i.e., level of statistical significance), and the partial correlation (i.e., the

relationship between each variable and percentage proficient when other variables are held constant).

The total correlation of all variables with percent reading proficient was .38. R^2 , a measure of the amount of variation shared among the predictor variables and percent proficient, was .14. Thus, the variables collectively explained 14% of the variation in percent reading proficient. This value is considered small (Cohen, 1988) and suggests that the student and teacher variables do not predict proficiency with a high degree of accuracy. Nevertheless, some variables are better predictors than others. Attendance rate and the percentage of teachers with a Master's Degree were statistically significant predictors. Districts in which attendance rates are higher and employ more teachers with advanced degrees tend to have higher proficiency rates. Percent ELL also explained a significant proportion of variance.

Table 4: Results of Regression analysis on STARS Reading Data

Variable	Mean	SD	Standardized Beta	t-test value	p value	Partial Correlation
ELL %	.018	.118	-.168	-2.17	.03	-.15
SPED %	.155	.045	-.041	-0.58	.56	-.04
FRL %	.348	.147	-.097	-1.15	.25	-.07
AR	.957	.018	.180	2.00	.04	.13
Graduation %	.897	.250	.012	0.17	.86	.01
Mobility %	.093	.072	.108	1.13	.26	.07
Accomd %	.050	.073	-.107	-1.49	.13	-.10
ATS	35576	3243	-.066	-0.76	.45	-.05
AYTE	16.48	2.77	.034	0.46	.65	.03
PTM	.287	.133	.187	2.28	.02	.15

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch %; AR = attendance rate; Accomd % = accommodation %; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 219 districts.

The data in Table 5 is based on 234 districts. All variables were entered into the regression equation at the same time to predict the percentage of students reaching proficiency on the STARS writing assessment. The total

correlation of all variables with percent writing proficient was .30. R^2 was .09. This value is considered small (Cohen, 1988) and suggests that the student and teacher variables do not predict proficiency with a high degree of accuracy. In addition, slightly less variation is explained in writing proficiency compared to reading proficiency. Two variables, Percent ELL and graduation rates, were statistically significant predictors. Note that the beta weight of Percent ELL is negative. This suggests that as the percentage of ELL learners increases, writing proficiency decreases. This outcome is consistent with the relationship between Percent ELL and reading proficiency displayed in Table 4.

Table 5: Results of Regression Analysis on STARS Writing Data

Variable	Mean	SD	Standardized Beta	t-test value	p value	Partial Correlation
ELL %	.020	.052	-.173	-2.21	.02	-.14
SPED %	.154	.044	-.088	-1.27	.20	-.08
FRL	.346	.146	-.063	-0.76	.44	-.05
AR	.956	.018	.118	1.32	.18	.09
Graduation %	.901	.243	.154	2.35	.02	.16
Mobility %	.097	.073	.049	0.51	.60	.04
Accomd %	.051	.072	.035	0.49	.64	.03
ATS	35779	3330	.017	0.19	.86	.01
AYTE	16.44	2.71	.029	0.39	.70	.03
PTM	.967	.136	.017	0.19	.84	.01

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; Accomd % = accommodation %; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 districts.

The results shown in Tables 4 and 5 indicate that variables have different effects on reading and writing proficiency. However, it is important to reiterate that the predictor variables explained little of the variation in proficiency levels. The relatively low amount of variation explained by district demographic characteristics suggests that data that is more closely tied to classroom practice and procedure may be necessary to better understand why some districts are outperforming others.

Question 4. We examined whether student and teacher variables were predictive of compliance or non-compliance with state and federal NCLB standards. A discriminant function analysis was conducted for students in the 4th, 8th, and 11th grades using the student and teacher variables described in Question 3. Two other variables were added, including number of students in each district and number of teachers in each district. We conducted separate analyses using state and federal compliance as outcomes. In the current analysis, we report the results of compliance with *state* standards rather than *federal standards* for two reasons. First, approximately half the districts in the database do not report on federal compliance. Thus, while 234 districts reported compliance/no compliance at the state level, 109 did so at the federal level. A second reason is that state and federal analyses mirror one another, although variables tend to be better discriminators using state compliance due to a much larger sample size (i.e., statistical power). The results using the federal data are available upon request.

Table 6 presents mean scores for the met versus not met groups for the three targeted grade levels. Table 7 shows corresponding F- and statistical significance values for the differences between the met versus not met groups at 4th, 8th, and 11th grades. Three variables consistently differ between the met versus not met groups. The most robust difference is on the attendance rate variable. Districts with higher attendance rates are more likely to meet state compliance. This pattern also occurred when comparing federal compliance. The most plausible explanation is that higher attendance leads to more instruction and learning; thus, making it more likely that districts will perform well. A second difference occurs with the average years of teaching experience variable. Districts with more experienced teachers are more likely to comply with state standards. However, it is unclear whether teaching experience has a direct causal effect on compliance rates, or whether teachers with more experience gravitate over time to districts with higher student performance and compliance rates (e.g., Carey, 2004; Darling-Hammond, 2000). A third difference occurred with the free and reduced lunch (FRL) variable. Districts with lower FRL percentages were more likely to comply with state standards.

Table 6: Means for Met Versus Not Met Status on State Compliance

	4 th Grade		8 th Grade		11 th Grade	
	Met	Not Met	Met	Not Met	Met	Not Met
ELL %	.01	.04	.02	.04	.02	.03
SPED %	.15	.16	.15	.16	.15	.16
FRL	.33	.36	.34	.42	.37	.42
AR	.96	.93	.96	.94	.96	.94
Graduation %	.87	.80	.89	.85	.96	.94
Mobility %	.10	.13	.09	.12	.09	.10
# Students	1172	558	1171	457	1132	444
ATS	35753	35394	35618	34790	35723	34456
AYTE	16.4	15.1	16.6	14.3	16.6	15.3
PTM	.30	.25	.29	.23	.30	.25
# Teachers	84.8	45.6	80.7	37.2	82.3	35.9

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; ATS = average teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 4th grade districts, 256 8th grade districts, and 253 11th grade districts.

Table 7: F Values and Significance for Met Versus Not Met Status on State Compliance

	4 th Grade		8 th Grade		11 th Grade	
	F	Sig.	F	Sig.	F	Sig.
ELL %	2.68	.10	5.40	.02	1.23	.27
SPED %	.41	.52	1.11	.29	2.41	.12
FRL	.56	.44	6.48	.01	6.61	.00
AR	20.61	.00	28.83	.00	20.25	.00
Graduation %	.84	.36	.62	.43	3.56	.06
Mobility %	2.72	.10	2.34	.13	.63	.43
# Students	.34	.56	.63	.43	.77	.38
ATS	.16	.69	1.21	.27	3.17	.08
AYTE	3.13	.08	14.51	.00	5.19	.02
PTM	1.82	.18	4.73	.03	2.62	.11
# Teachers	.31	.58	.58	.45	.73	.40

Note: ELL % = English language learner %; SPED % = special education %; FRL = free and reduced lunch; AR = attendance rate; ATS = average

teacher salary; AYTE = average years teaching experience; PTM = percentage of teachers with masters degree. Results based on 234 4th grade districts, 256 8th grade districts, and 253 11th grade districts.

The correlation between the set of discriminating variables and compliance status was .36 for the 4th grade data, .44 for 8th grade data, and .37 for the 11th grade data. All three of these correlations were statistically significant at $p < .01$. These findings indicate that discriminating variables behave quite consistently across the three grades levels and explain approximately 10% of the variation in compliance status. Although the 10% variation explained criterion is statistically significant, it is considered low because 90% of the variation in compliance status is unexplained. Thus, districts either reach compliance or fail to reach compliance based on a variety of variables not included in the present analyses.

Conclusions

The present analyses lead to several general conclusions regarding relationships between STARS and NRT assessments and district compliance with state and federal standards. These are summarized below.

1. Scores from STARS and NRT assessments are correlated within and between the 2001 and 2003 cohorts. High correlations are evident in Table 2 for 4th and 8th grade, whereas the correlation drops at the 11th grade. The lower correlation may be due to differences in classes taken in high school or more variability in assessments.
2. Adequate yearly progress is difficult to achieve (Linn, 2003). The Nebraska NRT data comparing 2001 and 2003 cohorts do not reveal any changes in performance between 2001 and 2003. In contrast, STARS assessments reveal modest but statistically significant progress across the cohorts. One explanation is that STARS assessments are better aligned to district and state standards, and therefore provide a more reliable and valid assessment of educational progress compared to NRTs.
3. Few of the variables used in these analyses were good predictors of proficiency or state and federal compliance standards. Only three variables, attendance rates, years of teacher experience, and percentage of teachers with Masters' Degrees, provided significant prediction across analyses. The

present findings suggest that increasing attendance rates may have an impact on percentage of students reaching proficiency and compliance with state and federal achievement standards. The NDE should consider ways to increase attendance, and perhaps further investigate which districts are at risk for low attendance rates. Increased numbers of teachers with long-term experience and advance training also was associated with proficiency and compliance rates. Experienced and more highly educated teachers may be better prepared to plan, implement, and assess learning due to practical experience and additional graduate training. In addition, it should also be noted that Percent ELL appeared to have an important effect on 8th grade district writing and reading proficiency.

4. Most of the variables used in the analyses of questions 3 and 4 were poor predictors of proficiency and compliance. *This should not be interpreted negatively.* For example, Tables 4-7 reveal that number of students and teachers in a district (i.e., district size) is unrelated to proficiency and compliance. Similarly, Percent SPED and Mobility Rates did not predict proficiency or compliance. These findings provide valuable information about variables that are not related to compliance.

Recommendations

Several recommendations seem reasonable in light of these analyses. We summarize these below.

1. Continue using STARS as a planning and assessment strategy. STARS appears to be a more sensitive measure of yearly progress than NRTs.
2. Promote attendance. In the present analyses, attendance rates were the best predictor of proficiency and compliance by a wide margin, especially among at-risk districts.
3. Recruit and utilize experienced teachers with graduate training. One way to do so is to use experienced teachers as mentors within their school or as in-service facilitators. A second way is to use experienced teachers to serve at-risk students.
4. Consider collecting data on additional variables not included in the present analyses. Variables that assess the quantity and quality of

instruction may be particularly important given the important role of attendance rates. Students who attend school more often are more likely to become proficient and meet state and federal standards. Currently, it is unclear what role the quantity (i.e., amount of time) and quality of instruction play.

References

Carey, K. (2004). The real value of teachers: Using new information about teacher effectiveness to close the achievement gap. *Thinking K-16*, 8 (1), 3-32.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives*, 8, no. 1, retrieved June 7, 2004 from <http://epaa.asu.edu/epaa/v8n1/>.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29, 4-16.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32, 3-13.

Little, R.J.A. & Rubin, D.B. (1987). Statistical analysis with missing data. New York: Wiley.

Marion, S., White, C., Carlson, D., Erpenbach, W. J., Rabinowitz, S. & Sheinker, J. (2002). *Making valid and reliable decisions in determining adequate yearly progress*. Washington, DC: Council of Chief State School Officers.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Schafer, J.L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.